

3 1761 116340159



Analytical Studies Branch

CA1
BS1
-1992
R49

Government
Publications

DEVELOPMENT OF LONGITUDINAL PANEL DATA FROM BUSINESS REGISTERS: Canadian Experience

by

John Baldwin, Richard Dupuy and William Penner

No. 49

Research Paper Series



Statistics
Canada

Statistique
Canada

Canada



ANALYTICAL STUDIES BRANCH RESEARCH PAPER SERIES

The Analytical Studies Branch Research Paper Series provides for the circulation, on a pre-publication basis, of research conducted by Branch staff, visiting Fellows and academic associates. The Research Paper Series is intended to stimulate discussion on a variety of topics including labour, business firm dynamics, pensions, agriculture, mortality, language, immigration, statistical computing and simulation. Readers of the series are encouraged to contact the authors with comments, criticisms and suggestions. A list of titles appears inside the back cover of this paper.

Papers in the series are distributed to Statistics Canada Regional Offices, provincial statistical focal points, research institutes, and specialty libraries. Each paper is catalogued on the DOBIS computer reference system and in various Canadian university library reference systems.

To obtain a collection of abstracts of the papers in the series and/or copies of individual papers (in French or English), please contact:

Publications Review Committee
Analytical Studies Branch, Statistics Canada
24th Floor, R.H. Coats Building
Ottawa, Ontario, K1A 0T6
(613) 951-8213

DEVELOPMENT OF LONGITUDINAL PANEL
DATA FROM BUSINESS REGISTERS:
Canadian Experience

by

John Baldwin, Richard Dupuy and William Penner

No. 49

Social and Economic Studies Division
Analytical Studies Branch
Statistics Canada
1992

This paper was originally presented at the 1992 International Conference on Business Survey Methods held in Montreal. It is forthcoming in the *International Journal of Economic Organization*.

Aussi disponible en français



Digitized by the Internet Archive
in 2023 with funding from
University of Toronto

<https://archive.org/details/31761116340159>

Development of Longitudinal Panel Data from Business Registers: Canadian Experience

by

John Baldwin, Richard Dupuy, and William Penner

Business and Labour Market Analysis Group and Business Register Division,
Statistics Canada, Ottawa, Ont. K1A OT6, Canada

Research at Statistics Canada has made extensive use of longitudinal data bases to study business demography. This paper describes the way in which the data required for these exercises have been created from Statistics Canada's Business Register. It recounts the problems that researchers faced when using the Business Register and the lessons learned. It then describes the manner in which the traditional problems of creating longitudinal identifiers was solved with the creative use of data that involved a labour-tracking exercise.

This paper was originally given to the 1991 International Roundtable on Business Survey Frames that was held in Australia. It is forthcoming in the Statistical Journal, UN Economic Commission for Europe.

Key Words: Longitudinal Data, Business Registers

1. Introduction

Information on the structure of businesses that can be obtained from Business Registers, when combined with other statistical data, offer considerable potential for the development of the longitudinal panels required for research into public policy issues. Traditionally, longitudinal analysis has not been the primary focus of business registers. Registers focus primarily on the provision of an accurate snap-shot of the business population at a point in time. However, with the addition of information that allows operating and legal entities to be linked over time, a business register becomes a powerful tool supporting the extension of the output of statistical agencies into the field of business demographics. This paper describes how information from business registers at Statistics Canada is being extracted and combined with other data in order to facilitate longitudinal analysis. Some of the resulting research studies are reported in the attached reference list.

Statistics Canada recently has supported several research initiatives that have examined the effect of trade liberalization on the industrial sector [2], that developed new indicators of the competitiveness of the industrial system [3,5], that outlined the effect of business growth and decline on conditions in labour markets [8], that allowed the dynamics of small, medium and large enterprises to be studied [4,9], and that evaluated the effect of mergers [7,10].

A large number of questions were investigated. They included: What was the birth and death rate in the manufacturing sector? How did it differ between foreign and domestic businesses? How has it responded to trade liberalization? Were Canadian businesses becoming more specialized as trade liberalization occurred? Did this process differ between multinational and foreign businesses? What was the growth path of large and small businesses? Were

productivity differences between different groups of businesses changing over time? Did job loss differ in small and large businesses and did it respond differentially to recessions? Did mergers serve to increase the productivity of acquired plants? Did foreign takeovers have a different impact than domestic takeovers? Do concentration measures capture the degree to which market share is being transferred from one group of businesses to another as a result of the competitive process?

These projects required the development of longitudinal panel data sets. Two longitudinal data sets have been generated for this purpose. The first covers the manufacturing sector. The development of the manufacturing panel data [6] and summaries of the results [1] were described at the Williamsburg International Roundtable on Business Survey Frames. (See U.S. Department of Commerce, Proceedings, International Roundtable on Business Survey Frames, Williamsburg, VA, October 9, 1990). The second data base covers most major sectors of the Canadian economy and was initially developed to produce measures of job turnover--the degree to which jobs are created in newly-identified and growing businesses and lost in no-longer-identified and declining businesses (See [11]).

The Canadian manufacturing sector was chosen initially as the focal point for most of the economic studies since the Census of Manufactures provided an ideal source for the creation of a longitudinal panel. The Canadian Census of Manufactures is taken at the plant or establishment level. It collects data about physical entities, mills, refinery plants, etc., irrespective of who owns them or how they relate to each other. Each establishment is assigned a unique identifier number that remains with it as long as the plant continues in the Census. This number does not change with ownership or name changes and, therefore, the appearance and disappearance of these

identifiers allows "real" births and deaths to be identified. In addition, each establishment is linked to an owning enterprise and thus commonly-owned establishments can be grouped together as businesses. Accordingly, longitudinal analysis can be done using either establishments or businesses.

The second longitudinal business data panel was developed for economic research on employment dynamics--part of a Longitudinal Employment Analysis Program (LEAP). It covers most sectors of the Canadian economy. It was constructed from administrative tax records that were not organized with longitudinal studies in mind and it has proven to be difficult to construct. It had to overcome the problems that are commonly associated with such files.

Many data bases that are constructed from administrative records without the benefit of supplementary data derived from surveys are unable to define the nature of births very precisely. Often this is because the identifier numbers that are attached to reporting units in these data sets change for administrative or operational reasons. Identifiers can change if a business moves from an unincorporated to an incorporated status. Or, identifiers can change if a merger occurs. In the latter case, merger entry and exit cannot be distinguished from greenfield entry and exit. Since the two have very different characteristics, failure to distinguish between them can produce misleading results.

The development of the LEAP data base has proceeded in two phases. In the first, the old Business Register system was used to create a longitudinal file. Many of the problems that arise in trying to build a longitudinal panel from a frame that was not meant for such an exercise occurred during the construction of LEAP. The second phase improved the longitudinal identifiers by linking data on the identity of employees to businesses. This phase illustrates the need for

creativity in editing administrative files if they are to be used for longitudinal studies.

This paper describes the manner in which problems in building LEAP panel data set were overcome. But first, a few comments are needed on several lessons that have been learned during the development and use of the Census of Manufactures panel data base about the characteristics of a central frame and associated data that facilitate longitudinal studies.

2. Some Requirements for the Development of Longitudinal Panels

In the course of preparing the longitudinal studies using the manufacturing panel data described above, several lessons emerged.

1) Flexibility is required in the definition of business units. The information derived from a business register and associated data often need to be recast to suit the needs of individual research projects. This means that sufficient characteristics relating to the structure of a business must be available so that researchers can choose the definition of a business that is most relevant for the research being undertaken. Research requires access to a business structure, the specification of signals that will be used to define units for analysis, and information on the business units that can be used with the signals.

2) A frame is useful only if a wide range of statistical data can be connected to it.

3) While units can be tracked longitudinally at various levels--the establishment or operating level, the legal entity level, or the top controlling entity level--being able to track business entities at the operating level is essential for a wide range of studies that need to define businesses narrowly at the industry level.

2.1. *Flexibility in Definition of a Business*

Many of the longitudinal research studies conducted at Statistics Canada examine different aspects of intra-industry firm dynamics. A study of firm dynamics normally requires that births and deaths be measured. There are many ways of defining births and deaths (see [4] for a paper that compares entry by plant creation and entry by merger). All have the defect of being arbitrary; but each can be used to shed light on a different aspect of analytical interest. Longitudinal studies require information that allows for alternative definitions of business births and deaths. This can either be information stored on the business register or it can be information that can be linked to the business register.

There are many definitions of a birth because a business is not defined in a single dimension. In general, it can only be described with a vector of characteristics. That, in and by itself, would be unimportant if only one of those characteristics were of interest when it came to defining births. This is not the case. Different questions require different definitions of a new business. Therefore, flexibility is required. Examples of the need for flexibility are:

- a) If we want to ask how the creation of new businesses affects employment, a greenfield definition of a new business is required--that is births are new businesses that offer new employment opportunities. This definition primarily depends upon the plant age and employment status of a business. New businesses are those that enter an industry by building plant not previously in existence.
- b) If we are interested in the effect of new business creation on the competitive process,

we need to measure several different concepts of newness. We need to know whether the new business adds to the stock of businesses in an industry (greenfield entry) or whether the identity of controlling interests has changed because acquisitions have brought new owners into an industry. Competition may be stimulated as much by a change in the identity of owners as by greenfield entry in situations where an oligopoly with mature participants had ceased to function competitively. Knowledge of when ownership changes take place is also important for studies that examine the effect of mergers.

c) In some cases, location will be required for the definition of entry--especially for the type of spatial studies done by geographers. A business may change nothing but its plant's location and be considered a new business in some situations but not in others. In competition studies, such a change would be important if the change introduced a new business into a regionally-fragmented market. This would not be classified as a birth (a new entity) for an industry with a national market. In a study that examined the regional impact of entry and exit, movement of a plant from one region to another would be classified as a birth.

Some research issues then relate to performance characteristics at the operating level of a business. For these, researchers will want to employ signals that relate to the status of plant to define a birth. Information on ownership, location, and age are required for studies at this level. Other researchers will want to investigate the effectiveness of different management types. In these cases, information is required about the ownership and legal constitution of the business. Other information, for example age and location, is important for demographic studies.

2.2. Comparability of Data through Frame Linkage

Many longitudinal firm studies evaluate the performance of different types of businesses (large versus small, domestic versus foreign) or businesses that change their circumstances (where ownership changes). Evaluations may involve the trade performance of domestic-controlled versus foreign-controlled businesses. They may track the investment activity of businesses in one region as opposed to another. They may examine the nature of the response of research and development expenditures to tax incentives. They may compare the financial results of businesses before and after mergers. Studies such as these require that businesses have a wide range of attributes attached to them. This requires that data from diverse sources be linked to business structures. Export and import data, investment data, financial data, and data on research and development have to be merged together in such a way that they are readily linked to business structure.

In order to accomplish this, the business register must be widely adopted in the statistical agency for frame purposes. A central register is not very useful for longitudinal studies if different versions are kept by the various survey divisions and if these versions cannot be easily linked.

2.3. Focal Point for Longitudinal Tracking--The Statistical Unit

Longitudinal studies may focus on operating, financial or legal structures. Studies that examine job creation will want to focus on the operating level and examine new establishments--both those associated with new businesses and those associated with existing businesses. This

requires the creation of a longitudinal identifier that tracks establishments over time.

Other longitudinal studies will want to group all plants within industries under common control--whether or not these make up less than a legal entity or several legal entities to examine how these units perform over time. For example, studies of competition need to define the business at the industry level. Once again this requires the creation of longitudinal identifiers for establishments. It also requires that establishments be linked together into a common ownership structure.

Another group of longitudinal studies will want to focus on the legal entity--where studies of profitability are contemplated--since this is generally the level at which profitability data are collected. These studies require a legal entity longitudinal identifier.

Finally, studies that focus on aggregate measures of concentration will want information on the ultimate controlling enterprise. For these studies, all businesses under common control will need to be considered jointly and longitudinal identifiers for the topmost controlling enterprise established.

Longitudinal studies then require that identifiers be produced that link various levels of a business together over time. Unfortunately, this is an arduous task. Priorities need to be established for profiling exercises when it is not feasible to create longitudinal identifiers at both the operating (establishment) and at the legal (business or company) level.

Business demographic studies are not the first to wrestle with these problems. Panel studies on income dynamics have had to decide whether to focus on tracking individual family units and to measure the component parts or to follow the individual components and create family units from them. Decisions to track the family rather than individuals have come to be

regretted since many of the policy issues relate to transitions of individuals from one unit to another. The same problems arise in business demographic studies if the business rather than its component parts are chosen for longitudinal analysis.

Several criteria can be used to establish whether, for research purposes, longitudinal identifiers should be created at the operating level, at the legal level, or at the top enterprise level. The first has to do with the facility with which rules can be derived and information obtained to create longitudinal identifiers. The second has to do with the range and value of studies that can be derived from each set of identifiers. Both considerations suggest that priority should be given to creating longitudinal identifiers at the establishment level.

First, on the basis of ease of definition, creating longitudinal establishment identifiers should have priority. Deciding whether an operating entity continues in existence can be done on the basis of location and product line. These are characteristics that are relatively easy to define and for which information is readily obtainable. Deciding whether a business continues in existence may be more difficult if ownership or control is used as the criterion for continuation. It is difficult to establish ownership and control because of the nuances given to these terms. Is control to be defined as ownership of more than 50% of equity or will lesser amounts suffice? Should just ownership of equity be taken into account or should convertible debt and preferred shares also be considered when defining control? Of course, the continuation of a legal entity might be chosen as the criterion for developing a longitudinal business file--but few interesting research questions can be answered with a file built on such a definition alone.

Second, the value of longitudinal identifiers at the operating level is greater than at the legal entity level in terms of resulting studies. In the Canadian case, where longitudinal identifiers

for establishment data were developed, most studies have built upwards from establishment data. Having a longitudinal identifier at this level was of critical importance. The research required that commonly-controlled establishments be aggregated to the industry level. Business data collected only at the legal entity level are too aggregated for many tasks.

The Canadian experience then in building business panel data for research from the Census of Manufactures suggests that the order of priorities for a research group that is concentrating on developing data for analytical projects should be:

- i) Establishing consistent rules to create continuity at the lowest level--the operating or establishment level.
- ii) Linking each establishment to a legal entity for which there is financial information.
- iii) Linking each legal entity to the top controlling enterprise level.
- iv) Establishing longitudinal links at the legal entity level or at the enterprise level--to capture changes in control or ownership.

None of this should be construed as requiring that business registers, as opposed to longitudinal panel data, be constructed in the above fashion. What is most important for longitudinal analysis may be inordinately costly for register maintenance. Automatic registration in administrative systems may exist only for legal entities and may dictate that registers focus at this level. Operating entities below the legal entity may be developed only as surveys use the legal frame and collect operating data. Development of longitudinal panels at the operating level will require that the signal to be used in defining a continuing business be clearly specified and

that the information that is required for implementation of the signal be derived from such surveys. For example, if a continuing establishment is to be defined as one that does not change ownership, location, name, or product line, then information on all these factors needs to be developed from surveys. In the end, the level at which a frame maintains continuity may not matter so much as the extent to which various characteristics of businesses--subcomponents corresponding to operating units--can be attached to the structure of businesses contained in the frame.

This suggests that two characteristics of business registers may suffice if longitudinal panels of business data are to be created. The first is the maintenance of continuity over time in terms of legal entities and the maintenance of links both downwards from the legal entities to operating divisions and upwards to parents. The second is a link to other sources of data that allow continuity rules to be developed at the level of operating entities or for the parent. Longitudinal files can then be generated for different purposes using specific signals--i.e., ownership changes at the level of the parent to signal changes in control, or employment data at the level of the establishment to signal births and deaths.

This does not mean that it is optimal to maintain all information that might be required for the creation of the longitudinal file outside the frame. Decentralization of data collection can lead to the emergence of gaps in information that have far reaching consequences elsewhere. Data that is essential for longitudinal analysis may not have much importance for survey divisions and may be the first to be omitted or the last to be edited if budgets tighten. Moreover, unless consideration is given to the way in which data available from specific surveys can be used for longitudinal panel creation, the required information may not be collected. By having the division

responsible for the frame focus on the characteristics that should be maintained in association with the central frame, these decisions may be better coordinated than if left to survey divisions.

Central frames should, therefore, be developed with consideration being given to storing some information with the frame that allows for the creation of longitudinal panels. Unfortunately, storage costs quickly mount as more and more information is collected. These costs must necessarily temper the desire to create a business register that is all things to all people. Perhaps the best we can manage is a register that maintains structure across as large a sample of the business population as possible and is as current as possible. Ambitious new data bases will still have to be creative in their use of associated data to develop longitudinal panels necessary to address different analytical issues.

In what follows, we describe the development of one such longitudinal panel at Statistics Canada. The remainder of the paper describes how external data were used in a creative fashion to construct a longitudinal file to examine the job-creation and job-destruction process in Canadian businesses.

3. The Longitudinal Employment Analysis Program

The Longitudinal Employment Analysis Program (LEAP) was designed to provide longitudinal data on the behaviour of employment levels of Canadian businesses. The data base makes use of administrative tax records, data from the Business Register and from surveys on average wage rates to derive the employment profile of businesses over time. Industry level data are produced on the employment in businesses that appear for the first time, that disappear, that

grow and that decline. The data have been used to investigate the dynamics of job growth and decline.

The research program focused on estimating the amount of job turnover due to the dynamics of business growth and decline. An example of the output is presented in Table 1. In this table, the business population is divided into newly-identified, no-longer-identified, growing and declining continuously identified businesses. Employment (ALU's) in each of these categories is provided for 1978 and 1984. Between 1978 and 1984, the business count grew from 601,448 to 782,196. The net increase of 180,748 resulted from 443,914 births and 263,166 deaths. There were 1,539,700 jobs in 1984 in businesses that were newly-identified since 1978--16.95% of base-year employment. There were 1,208,400 jobs in 1978 located in businesses that were no-longer-identified by 1984--13.2% of 1978 employment. Growing incumbents added 1,692,000 jobs--18.5% of 1978 employment--over the period; declining businesses lost 1,106,000 jobs--12.1% of base-year employment.

The employment record of each business is derived from administrative taxation records that each Canadian employer must file. These taxation data include, *inter alia*, gross earnings for each employee for the calendar year. They cover each individual who earned \$500 or more from any single employer in any year and extend from 1978 to 1989.

The payroll data that are filed by employers with Revenue Canada is associated with a Revenue Canada employer identification number--a payroll deduction account number (PAYDAC or PD number). Employers may have more than one payroll number. Statistic Canada's Business Register assigns all businesses a Business Register Identification (BRID) number and links this business number to payroll numbers.

Using this link, the longitudinal employment analysis programme aggregates payroll data to the business level to create total earnings of employees in a business. Employment data then are derived by dividing a business's annual payroll by estimates of annual average earnings derived from survey material. The resulting employment statistic is referred to as an average labour unit (ALU).

The longitudinal employment data base was developed using the old Business Register system. An employer registration form received by Statistics Canada from Revenue Canada provided the major source of information that was used to derive the universe of employers and the industry classification for each business. New payroll numbers that were associated with new businesses were classified as newly-identified businesses.

Different events could generate a new business. The incorporation of previously unincorporated businesses could trigger the death of one business identification number and the issuance of a new one. The amalgamation of two or more corporations into one corporation could cause the original business identification number to disappear and a new one to emerge. A spinoff of part of a company to new controlling interests could cause the birth of a new business number.

The Business Register tried to develop some continuity in the assignment of business numbers. In the case of incorporations, the Business Register tried to match the "old" and the "new" business numbers. In the case of control changes, provision was made for the storage of information that allowed the business number that died as the result of amalgamation, spin-off, or other control change to be linked to the new business number. But the primary function of this register was to provide, in a timely manner, the statistical sampling frame required by Statistic

Canada's business surveys. Building in all the events required for longitudinal analysis and maintaining information on these events was not given priority.

The longitudinal employment analysis program was interested in the first instance in relating the dynamics of the business sector--business growth and decline--to employment change. Employment change was of interest primarily, though not exclusively, because of an interest in job creation. Therefore, the group that focused on the creation of the longitudinal business identifier for the employment analysis programme focused on modifying the existing frame so that longitudinal business identifiers could be used to measure job change. This meant that the classification--"newly-identified"--was designed to contain primarily births that created new jobs. The classification--"no-longer-identified"--was meant to contain deaths where jobs disappeared. False births and deaths--where the appearance and disappearance of a business identification number was not as likely to be related to job creation and destruction--had to be eliminated.

The first version of the longitudinal file used in the longitudinal analysis programme accomplished this by examining all large births and deaths manually. Public lists of mergers were used to check that amalgamation had not resulted in births and deaths. Despite the care that was taken in editing the file manually, the extent to which jobs appearing in newly-identified units could be associated with job creation was uncertain because the accuracy of the editing process that created the longitudinal files had not been assessed. This was done in two stages.

In the first stage, a computerized procedure checked the previous editing system which relied on expert judgement. A program was used to match all births to all deaths and vice versa--using names and payroll account numbers. Newly-identified-businesses were matched to no-longer-identified and continuing businesses. A similar procedure was followed for no-longer-

identified businesses. The resulting matches were then carefully examined using Business Register information and error rates were calculated. At an aggregate level, the error rates from misclassified births and deaths were small--about 10%. However, error rates were much larger for some industries and many research studies need accurate data at the industry level.

While name and payroll number matching programmes increased the degree to which the validation process could be computerized, it still required too much manual intervention. Moreover, it only caught those situations where business identifier numbers were dropped but a similar name was retained. A second procedure was, therefore, developed to improve and to automate the routine that checked for false births and deaths.

This procedure relied on tracking the workforce of businesses over time. The data base used to create the longitudinal file relied on tax remittance data filed by businesses regarding employee remuneration and contained social insurance numbers for all employees. This meant that employees could be followed from one payroll account to another and the percentage of workers present in a business in one year that could be found in another could be calculated. Labour was, therefore, tracked from one firm to another and this path was used to establish whether a business was potentially linked to another.

The editing programme that was then chosen used both automated name matching and labour tracking. Labour tracking was first used to narrow the scope and, therefore, the cost of name-match routines. A list of potential links was created by finding all businesses that shared a worker with either a birth or death. This created a set of businesses with which the birth or death had a strong link. Then births (deaths) were matched against this potential set for name similarities and for the occurrence of common payroll account numbers. It was discovered that

these matches generally were either false births or deaths.

Using name-match routines only on the sample of businesses that shared employees rather than name-matching procedures on the entire sample provided a substantial improvement for two reasons. First, the 100,000 odd births and the 100,000 odd deaths that the annual version of the administrative file contained before editing would have had to be matched against over 600,000 continuing businesses. By reducing the number of searches, we both decreased computer costs and at the same time were able to use more sophisticated and costly matching routines. Second, name-matching routines often turn up partial or imperfect matches which then have to be carefully checked manually. By reducing the exercise to just those businesses which had a strong labour link, the manual intervention phase was eliminated--since research showed that cases where there was an imperfect name match, but where labour force was shared, almost always involved false births or deaths.

Labour force tracking was then used to delineate an additional set of cases that turned out invariably to be false births and deaths. If a business was falsely identified as dying, a substantial number of its workers should be found in another unit the following year. To check how often this occurred, a profile of worker continuation rates--the percentage of a business's workers that were found in another business in the subsequent year--was developed, both for continuing businesses and for no-longer-identified businesses--deaths. The business with the largest such figure was chosen as the target business with which the originating business unit had the closest affiliation--the dominant link. The percentage of shared workers (referred to here as the pass-through rate) with the dominant link was calculated. Figures 1 and 2 present a summary of these worker pass-through rates for a sample of large businesses in 1987. Figure 1 presents a simple

frequency distribution of the percentage of the sample that experience different pass-through rates. For deaths, the pass-through rates are concentrated in the 20-40% range. Pass-through rates for continuing businesses are concentrated above 60%. Figure 2 presents the cumulative frequency distribution of pass-through rates--where deaths are cumulated from 0% and continuing businesses are cumulated from 100%. It is evident that for this sample less than 15% of deaths had pass-through rates of 75% or more. Pass-through rates of 75% were symptomatic that the unit was a continuing business. The profile for continuing businesses and for deaths is quite different and this fact was used to aid in the second phase of the editing programme.

The difficulty in translating this observation into a workable rule to identify a false death arose because the cumulative distribution of the pass-through curve for no-longer-identified businesses differed substantially by size class. For businesses with more than 10 employees, it was discovered that few deaths had pass-through rates of 75% or more whereas most continuing businesses had pass-through rates at least this high. Therefore, when no-longer-identified businesses that had more than 10 employees had pass-through rates of 75% or more, these cases were reclassified as continuing businesses. Similarly, where more than 75% of a birth came from another business, this was classified as a false birth. A more restrictive rule was employed for smaller businesses since there was less difference in pass-through rates between small continuing and no-longer-identified businesses. It was found that in very small businesses (those between 5 and 9 employees), only cases where 100% of employees could be found in one business in the following year were invariably false deaths.

To summarize, no-longer-identified businesses in the administrative file were reclassified as continuing if there was a dominant employee link to a birth or to a continuing firm and

- 1) a perfect name match
- 2) a partial name match
- 3) a match with a payroll account number
- 4) a high proportion of the employees of the no-longer-identified business were all found in another firm--the proportion differing for small and large businesses.

Rule #4 was devised after manual evaluation. Whether it was too restrictive can be evaluated by comparing the average pass-through rate of these reclassified businesses compared to the pass-through rates generated by the mechanical matching rules #1 to #3. Table 2 contains these rates both for no-longer identified businesses that were linked to newly-identified businesses and to continuing businesses for the year 1989-90. The average continuing business had a pass-through rate of 73%. The average no-longer identified business had only a 49% pass-through rate. The no-longer-identified businesses that were reclassified as continuing had pass-through rates that were close to the continuing business population. Interestingly, case 4 had about the same pass-through rate as the other cases--indicating that the arbitrary rule that was adopted was probably quite accurate and that further fine-tuning was not justified.

The pass-through data and related information can be used to provide information that will complement the job-growth and job-decline data that the project initially had set out to provide. The pass-through rate of continuers--some 73%-- and no-longer-identified businesses--some 49%--provides an indication of the extent of worker mobility in different businesses. Other information can be generated on the percentage of all workers who are found in the same industry or in other industries.

4. Conclusion

A business register that permits longitudinal panel data to be developed will facilitate economic research on business units. This greatly extends the range of services that a statistical agency can offer its clients. Business registers have not been developed with this objective in mind. But they can be utilized to provide histories of business structures with careful planning.

The history of the development of the LEAP file demonstrates the benefits of providing this service. It also demonstrates that substantial development work for the creation of longitudinal research files is required. This is likely to be the case for other development exercises. Most research projects are idiosyncratic and require definitions of business entities that are specific to the research agenda being undertaken.

Acknowledgements

The authors are indebted to John McVey, Andre Monty, Jacob Ryten and Garnett Picot for comments.

References

- [1] J. R. Baldwin (1990). The Dynamics of Firm Turnover and the Competitive Process. A paper prepared for the Fifth International Roundtable on Business Survey Frames, held in Williamsburg Va, 1990.
- [2] J. R. Baldwin & P. K. Gorecki (1986a). The Role of Scale in Canada/U.S. Productivity Differences in the Canadian Manufacturing Sector in the 1970s. Volume 6 of the research series of the Royal Commission on the Economic Union and Development Prospects for Canada. University of Toronto Press, Toronto.
- [3] J. R. Baldwin & P. K. Gorecki (1986b). "International Trade, Secondary Output and Concentration in Canadian Manufacturing Industries". *Applied Economics*. 18 (May): pp. 529-43.
- [4] J. R. Baldwin & P. K. Gorecki (1987a). "Plant Creation Versus Plant Acquisition: The Entry

Process in Canadian Manufacturing". *International Journal of Industrial Organization* 5: pp. 27-41.

[5] J. R. Baldwin & P. K. Gorecki (1989a). "Measuring the Dynamics of Market Structure". *Annales d'Economie et de Statistique* 15/16 (July/December): pp. 316-32.

[6] J. R. Baldwin & P. K. Gorecki (1990a). "Measuring Entry and Exit to the Canadian Manufacturing Sector Using Longitudinal Data: Methodology". In A.C. Singh (ed). *Analysis of Data in Time*. Proceedings of a Conference sponsored by Statistics Canada, Carleton and Ottawa University. pp. 255-70.

[7] J. R. Baldwin & P. K. Gorecki (1990b). "Mergers Placed in the Context of Firm Turnover". Proceedings of the Census Bureau Fifth Annual Research Conference. Bureau of the Census, Washington, D.C. pp. 53-73.

[8] J. R. Baldwin & P. K. Gorecki (1990c). *Structural Change and the Adjustment Process: Perspectives on Firm Growth and Worker Turnover*. Economic Council of Canada, Ottawa.

[9] J. R. Baldwin & P. K. Gorecki (1991a). "Firm Entry and Exit in the Canadian Manufacturing Sector". *Canadian Journal of Economics*. (May): pp. 300-23.

[10] J. R. Baldwin & R. E. Caves (1991b). "Foreign Multinational Enterprises and Merger Activity in Canada". In L. Waverman (ed.) *Corporate Globalization through Mergers and Acquisitions*. Investment Canada and University of Calgary Press, Calgary. pp. 89-122.

[11] Statistics Canada (1988). *Developing a Longitudinal Database on Businesses in the Canadian Economy*. Catalogue #18-501E. Minister of Supply and Services, Ottawa.

Biographical Information

John Baldwin is an advisor on economic research in the Business and Labour Market Analysis Group of Statistics Canada and is associated with the Canadian Centre for Management Development in Ottawa. He received his B.A. in economics from Queen's University in Kingston, Ontario and his Ph.D. from Harvard University. His interests lie in business demographics and industrial economics.

Richard Dupuy is a research analyst in the Business and Labour Market Analysis Division of Statistics Canada. He received his B.A. in economics from the University of Ottawa. His interests lie in longitudinal analysis and data base development.

William Penner is an economist working in the Business Register Division of Statistics Canada. He received his B.A. in public administration from the University of Saskatchewan. He is involved in the research associated with the business register at Statistics Canada.

Table 1. ALU Employment Change Firm Type, Canada, 1978-84

Firm Category	1978		1984		net change
	businesses	ALUs (000)	businesses	ALUs (000)	ALUs
Total	601,448	9126.5	782,196	10,043.3	916.8
Continuously Identified	338,282	7918.2	338,282	8,503.7	585.5
- increasing ALUs	189,094	4042.2	189,094	5,734.4	1,692.1
- decreasing ALUs	149,188	3875.9	149,188	2,769.3	-1,106.6
Previously - Identified			448,914	1,539.7	1,539.7
No Longer - Identified	263,166	1208.4	—	—	-1,208.4

Sources: Statistics Canada. [11]

Table 2. Worker Pass - Through Rates, 1989-90

	Pass-Through Rates (%)
Mislabelled No-Longer - Identified units	
a) linked to births	
(i) Case 1	67.2
(ii) Case 2	69.7
(iii) Case 3	69.1
(iv) Case 4	70.7
b) linked to continuing entities	
(i) Case 1	70.5
(ii) Case 2	66.9
(iii) Case 3	69.7
(iv) Case 4	73.0
Continuing Units	72.3
No-Longer - Identified Units	49.1

Note: For definitions see text

Source: Business and Labour Market Analysis, Statistics Canada

Distribution of Firms by Worker Pass-through rates: Continuing versus Deaths

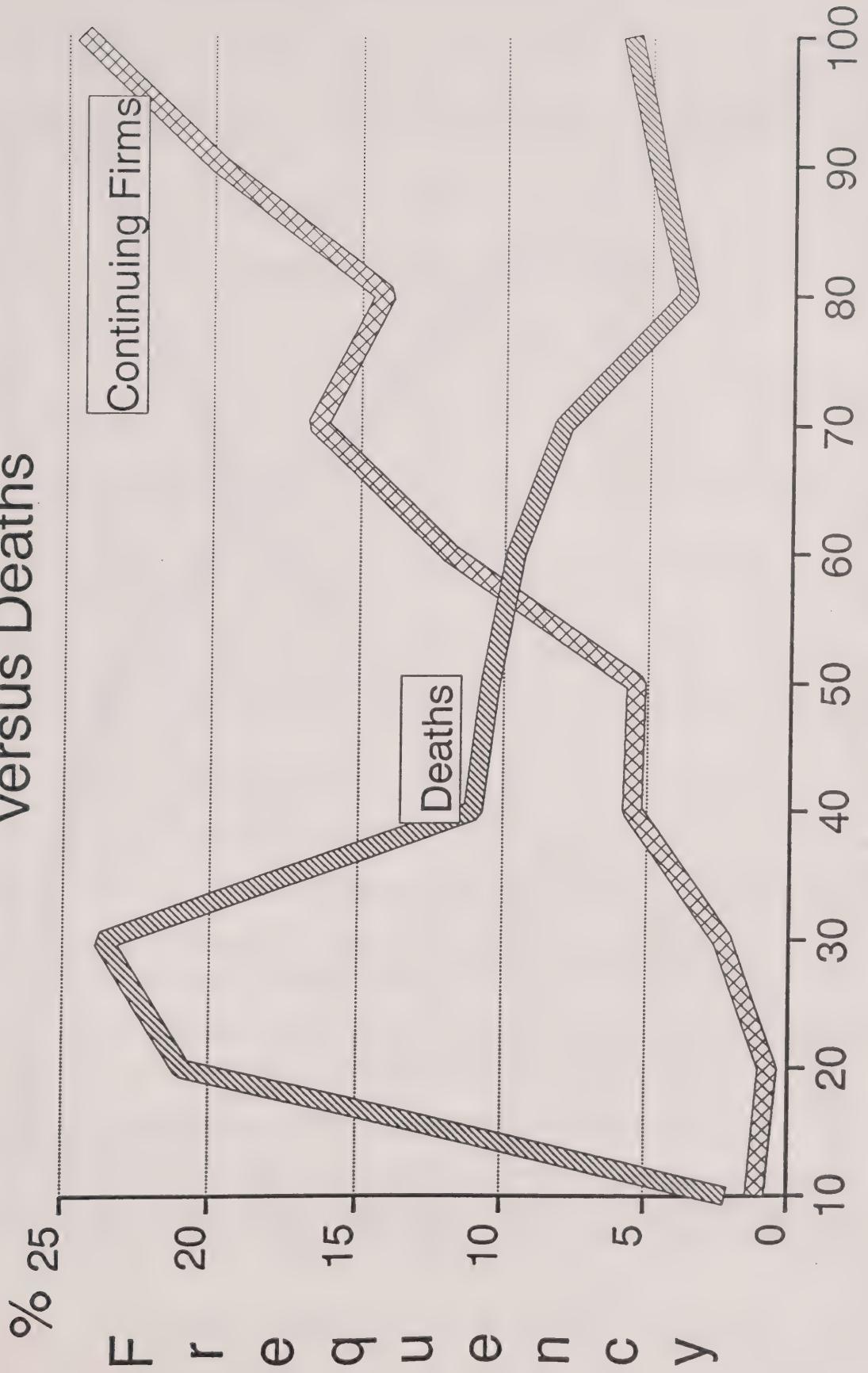


Figure 1

Figure 1

The Cumulative Distribution of Firms by Percentage of Worker Pass-through

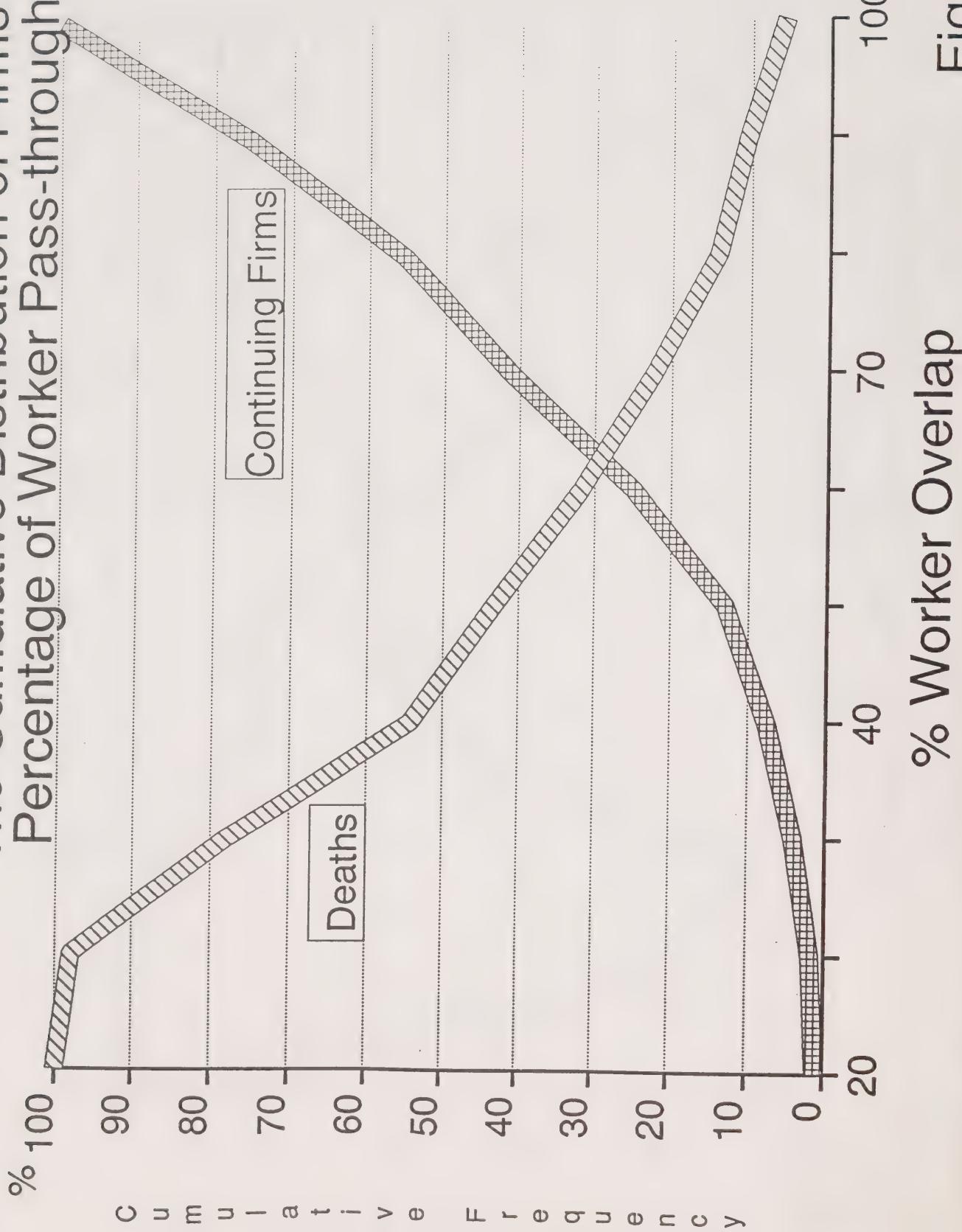


Figure 2
% Worker Overlap

ANALYTICAL STUDIES BRANCH
RESEARCH PAPER SERIES

No.

1. *Behavioural Response in the Context of Socio-Economic Microanalytic Simulation, Lars Osberg*
2. *Unemployment and Training, Garnett Picot*
3. *Homemaker Pensions and Lifetime Redistribution, Michael Wolfson*
4. *Modelling the Lifetime Employment Patterns of Canadians, Garnett Picot*
5. *Job Loss and Labour Market Adjustment in the Canadian Economy, Garnett Picot and Ted Wannell*
6. *A System of Health Statistics: Toward a New Conceptual Framework for Integrating Health Data, Michael C. Wolfson*
7. *A Prototype Micro-Macro Link for the Canadian Household Sector, Hans J. Adler and Michael C. Wolfson*
8. *Notes on Corporate Concentration and Canada's Income Tax, Michael C. Wolfson*
9. *The Expanding Middle: Some Canadian Evidence on the Deskilling Debate, John Myles*
10. *The Rise of the Conglomerate Economy, Jorge Niosi*
11. *Energy Analysis of Canadian External Trade: 1971 and 1976, K.E. Hamilton*
12. *Net and Gross Rates of Land Concentration, Ray D. Bollman and Philip Ehrensaft*
13. *Cause-Deleted Life Tables for Canada (1972 to 1981): An Approach Towards Analyzing Epidemiologic Transition, Dhruva Nagnur and Michael Nagrodska*
14. *The Distribution of the Frequency of Occurrence of Nucleotide Subsequences, Based on Their Overlap Capability, Jane F. Gentleman and Ronald C. Mullin*
15. *Immigration and the Ethnolinguistic Character of Canada and Quebec, Réjean Lachapelle*
16. *Integration of Canadian Farm and Off-Farm Markets and the Off-Farm Work of Women, Men and Children, Ray D. Bollman and Pamela Smith*

17. *Wages and Jobs in the 1980s: Changing Youth Wages and the Declining Middle*, *J. Myles, G. Picot and T. Wannell*
18. *A Profile of Farmers with Computers*, *Ray D. Bollman*
19. *Mortality Risk Distributions: A Life Table Analysis*, *Geoff Rowe*
20. *Industrial Classification in the Canadian Census of Manufactures: Automated Verification Using Product Data*, *John S. Crysdale*
21. *Consumption, Income and Retirement*, *A.L. Robb and J.B. Burbridge*
22. *Job Turnover in Canada's Manufacturing Sector*, *John R. Baldwin and Paul K. Gorecki*
23. *Series on The Dynamics of the Competitive Process*, *John R. Baldwin and Paul K. Gorecki*
 - A. *Firm Entry and Exit Within the Canadian Manufacturing Sector*.
 - B. *Intra-Industry Mobility in the Canadian Manufacturing Sector*.
 - C. *Measuring Entry and Exit in Canadian Manufacturing: Methodology*.
 - D. *The Contribution of the Competitive Process to Productivity Growth: The Role of Firm and Plant Turnover*.
 - E. *Mergers and the Competitive Process*.
 - F. *(in preparation)*
 - G. *Concentration Statistics as Predictors of the Intensity of Competition*.
 - H. *The Relationship Between Mobility and Concentration for the Canadian Manufacturing Sector*.
24. *Mainframe SAS Enhancements in Support of Exploratory Data Analysis*, *Richard Johnson and Jane F. Gentleman*
25. *Dimensions of Labour Market Change in Canada: Intersectoral Shifts, Job and Worker Turnover*, *John R. Baldwin and Paul K. Gorecki*
26. *The Persistent Gap: Exploring the Earnings Differential Between Recent Male and Female Postsecondary Graduates*, *Ted Wannell*
27. *Estimating Agricultural Soil Erosion Losses From Census of Agriculture Crop Coverage Data*, *Douglas F. Trant*
28. *Good Jobs/Bad Jobs and the Declining Middle: 1967-1986*, *Garnett Picot, John Myles, Ted Wannell*
29. *Longitudinal Career Data for Selected Cohorts of Men and Women in the Public Service, 1978-1987*, *Garnett Picot and Ted Wannell*

30. *Earnings and Death - Effects Over a Quarter Century, Michael Wolfson, Geoff Rowe, Jane F. Gentleman and Monica Tomiak*

31. *Firm Response to Price Uncertainty: Tripartite Stabilization and the Western Canadian Cattle Industry, Theodore M. Horbulyk*

32. *Smoothing Procedures for Simulated Longitudinal Microdata, Jane F. Gentleman, Dale Robertson and Monica Tomiak*

33. *Patterns of Canadian Foreign Direct Investment Abroad, Paul K. Gorecki*

34. *POHEM - A New Approach to the Estimation of Health Status Adjusted Life Expectancy, Michael C. Wolfson*

35. *Canadian Jobs and Firm Size: Do Smaller Firms Pay Less?, René Morissette*

36. *Distinguishing Characteristics of Foreign High Technology Acquisitions in Canada's Manufacturing Sector, John R. Baldwin and Paul K. Gorecki*

37. *Industry Efficiency and Plant Turnover in the Canadian Manufacturing Sector, John R. Baldwin*

38. *When the Baby Boom Grows Old: Impacts on Canada's Public Sector, Brian B. Murphy and Michael C. Wolfson*

39. *Trends in the distribution of Employment by Employer Size: Recent Canadian Evidence, Ted Wannell*

40. *Small Communities in Atlantic Canada: Their Industrial Structure and Labour Market conditions in the Early 1980s, Garnett Picot and John Heath*

41. *The Distribution of Federal/Provincial Taxes and Transfers in rural Canada, Brian B. Murphy*

42. *Foreign Multinational Enterprises and Merger Activity in Canada, John Baldwin and Richard Caves*

43. *Repeat Users of the Unemployment Insurance Program, Miles Corak*

44. *POHEM -- A Framework for Understanding and Modelling the Health of Human Population, Michael C. Wolfson*

45. *A Review of Models of Population Health Expectancy: A Micro-Simulation Perspective, Michael C. Wolfson and Kenneth G. Manton*

46. *Career Earnings and Death: A Longitudinal Analysis of Older Canadian Men*, Michael C. Wolfson, Geoff Rowe, Jane Gentleman and Monica Tomiak
47. *Longitudinal Patterns in the Duration of Unemployment Insurance Claims in Canada*, Miles Corak
48. *The Dynamics of Firm Turnover and the Competitive Process*, John Baldwin
49. *Development of Longitudinal Panel Data from Business Registers: Canadian Experience*, John Baldwin, Richard Dupuy and William Penner
50. *The Calculation of Health-Adjusted Life Expectancy for a Multi-Attribute Utility Function: A First Attempt*, J.-M. Berthelot, R. Roberge and M.C. Wolfson
51. *Testing The Robustness of Entry Barriers*, J. R. Baldwin, M. Rafiquzzaman
52. *Canada's Multinationals: Their Characteristics and Determinants*, Paul K. Gorecki

For further information, contact the Chairperson, Publications Review Committee, Analytical Studies Branch, R.H. Coats Bldg., 24th Floor, Statistics Canada, Tunney's Pasture, Ottawa, Ontario, K1A 0T6, (613) 951-8213.

